

HIV-1 and HIV-2 Sequence Subtypes

1. Introduction

In addition to the six HIV1 sequence subtypes A through F recognized in the 1993 database compendium (III-2 of APR 94), there are now two new subtypes to be considered, G and H, for what is now being called the M group of HIV-1 sequences (see Janssens et al., *AIDS Res. Hum. Retroviruses* **10**:877–879, 1994). Furthermore, we are aware of several sequences that don't fit at this time into any of the defined sequence subtypes of group M. And, as the cover of this compendium reveals, the O group of HIV-1s includes sequences that are as different from one another as M group subtypes. (The O group CA9 and VI686 sequences have not been made public, but were kindly provided to us for the use in the cover figure by the Institute of Tropical Medicine, Antwerp.) Together, the M and O group sequences of HIV-1s appear to constitute a “double star” phylogeny (Myers, *AIDS Res. Hum. Retroviruses* **10**:1317–1322, 1994).

Heretofore, the database has recognized two subtypes of HIV-2s, A and B. A recent study by Hahn's group at the University of Alabama argues for the existence of at least five HIV-2 sequence subtypes, A through E (Gao et al., *J. Virol.* **68**:7433–7447, 1994; the criteria used were virtually identical to those used for defining HIV-1 group M sequence subtypes).

Both HIV-1 and HIV-2 data sets present some evidence for hybrid viruses (see Sharp et al., *AIDS* **8** (suppl 1):S27–S42, 1994).

2. HIV-1 Phylogenetic Trees

The availability of many new HIV-1 *env* gp120 and gp160 coding sequences (Part I) permits us to generate new phylogenetic trees that can be compared to i) HIV-1 *gag* cds trees and ii) partial *env* cds trees. Since the publication of the 1993 compendium, we have been experimenting with asymmetrically-weighted parsimony analysis (as implemented by Hillis et al., *Science* **264**:671–677, 1994) as a way to reduce effects of homoplasy. The first step in this process is to run ordinary parsimony analysis in order to determine the substitution biases; all things being equal, the biases will tend to follow the base composition of the sequences. We use Macintosh versions of PAUP and MacClade to accomplish this step. A resulting substitution frequency matrix (for HIV-1 *env* gp120 sequences) is shown below:

to:	A	C	G	T
from:	A	0.104	0.222	0.074
	C	0.060	0.023	0.095
	G	0.163	0.027	0.022
	T	0.052	0.112	0.046

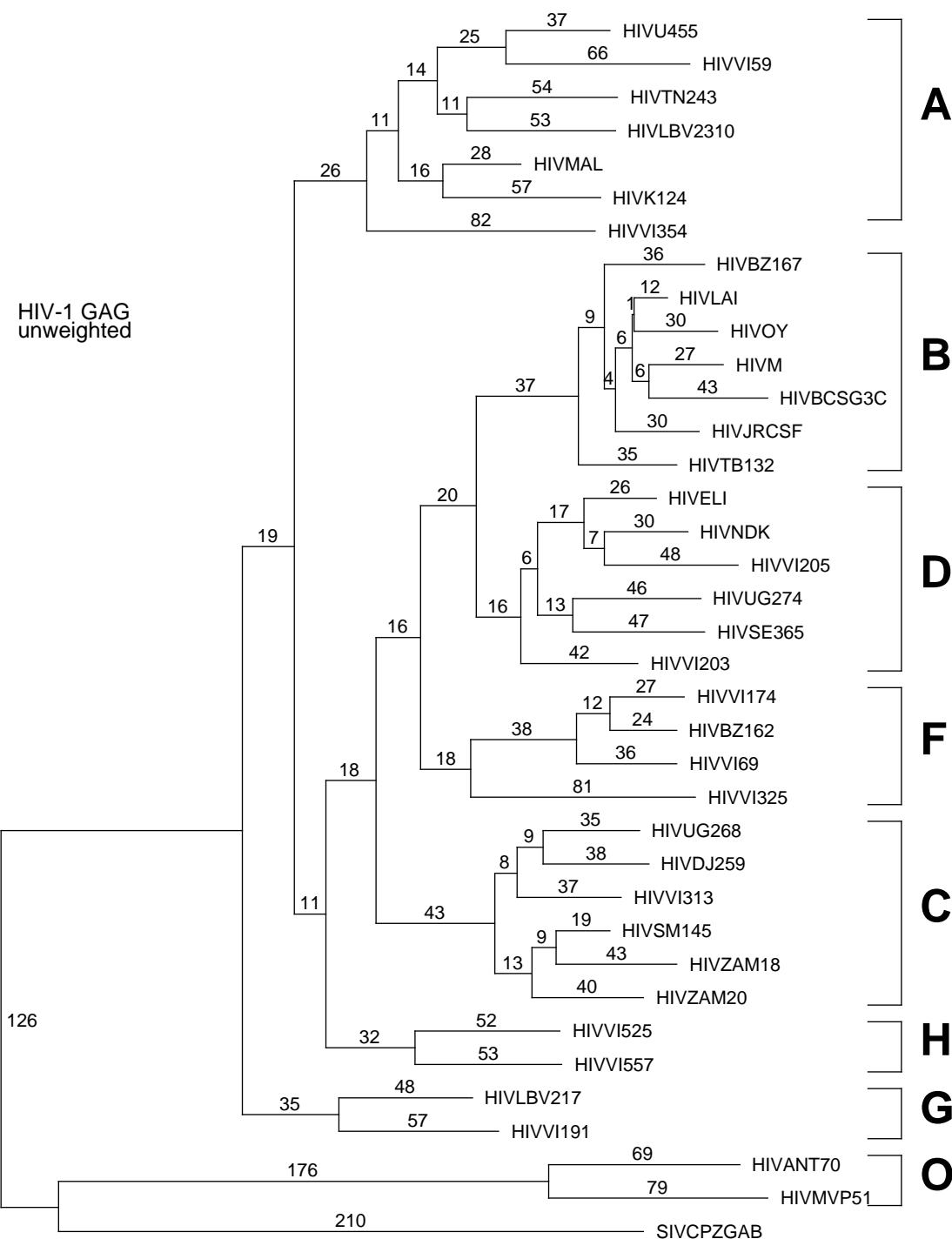
We see from this table that some transversions (e.g., A to C) are as frequent as some transitions (e.g., C to T). Purine-purine transitions occur at almost twice the frequency of pyrimidine-pyrimidine transitions. These appear to be consequences of the unusual base composition of HIVs and SIVs.

The next step in the analysis is to re-execute the parsimony algorithm (PAUP) using an inverse weighting table generated from the substitution matrix. Hence, the least common changes (G to T) are given the greatest weight and the most common substitutions (A to G) are given the least weight. In applying the inverse weighting rule, it may be necessary to truncate values in order that the triangle inequality relationship be preserved (see Maddison and Maddison: MacClade:Analysis of Phylogeny and Character Evolution. Sinauer Associates, Sunderland MA, 1992). Because the branch lengths in the newly generated tree are derived from weighted terms, they are no longer indicative of the minimal number of base changes, as they would be in ordinary parsimony. However, the lengths do provide accurate relative distances. We continue to employ maximum likelihood analysis, neighbor-joining, and bootstrapping in order to strengthen inferences based upon parsimony tree analyses.

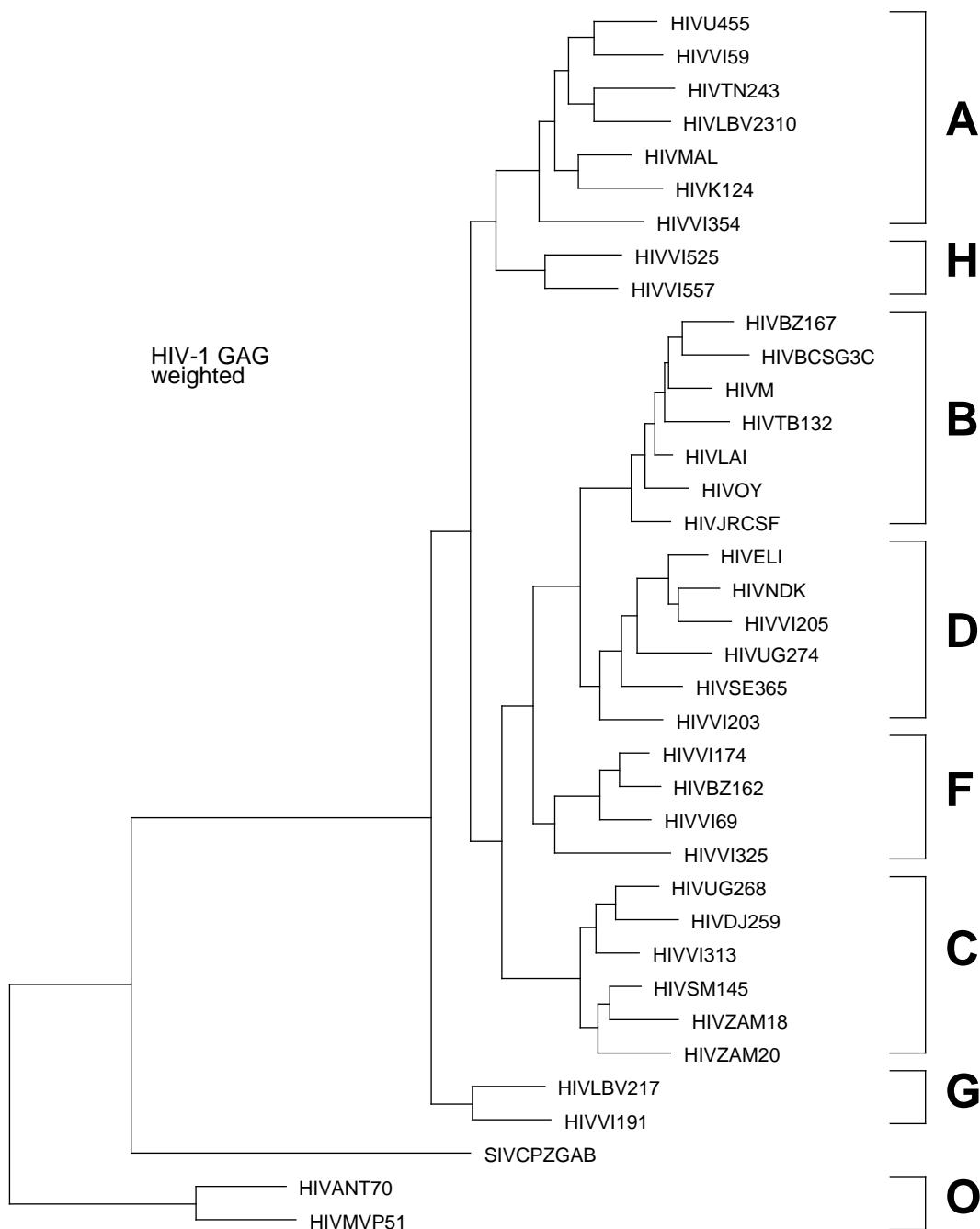
When the signal to noise ratio is sufficiently high, the weighted approach won't drastically change the tree analysis; however, it will invariably change one or two individual relationships, attachments to a major trunk. Weighted parsimony is especially warranted when analyzing “deeper” phylogenetic relationships that have attained “mutational saturation” (III.D).

In the following analyses, both unweighted and weighted trees are shown for HIV-1 *gag*, *env* and *tat* coding sequences in that order.

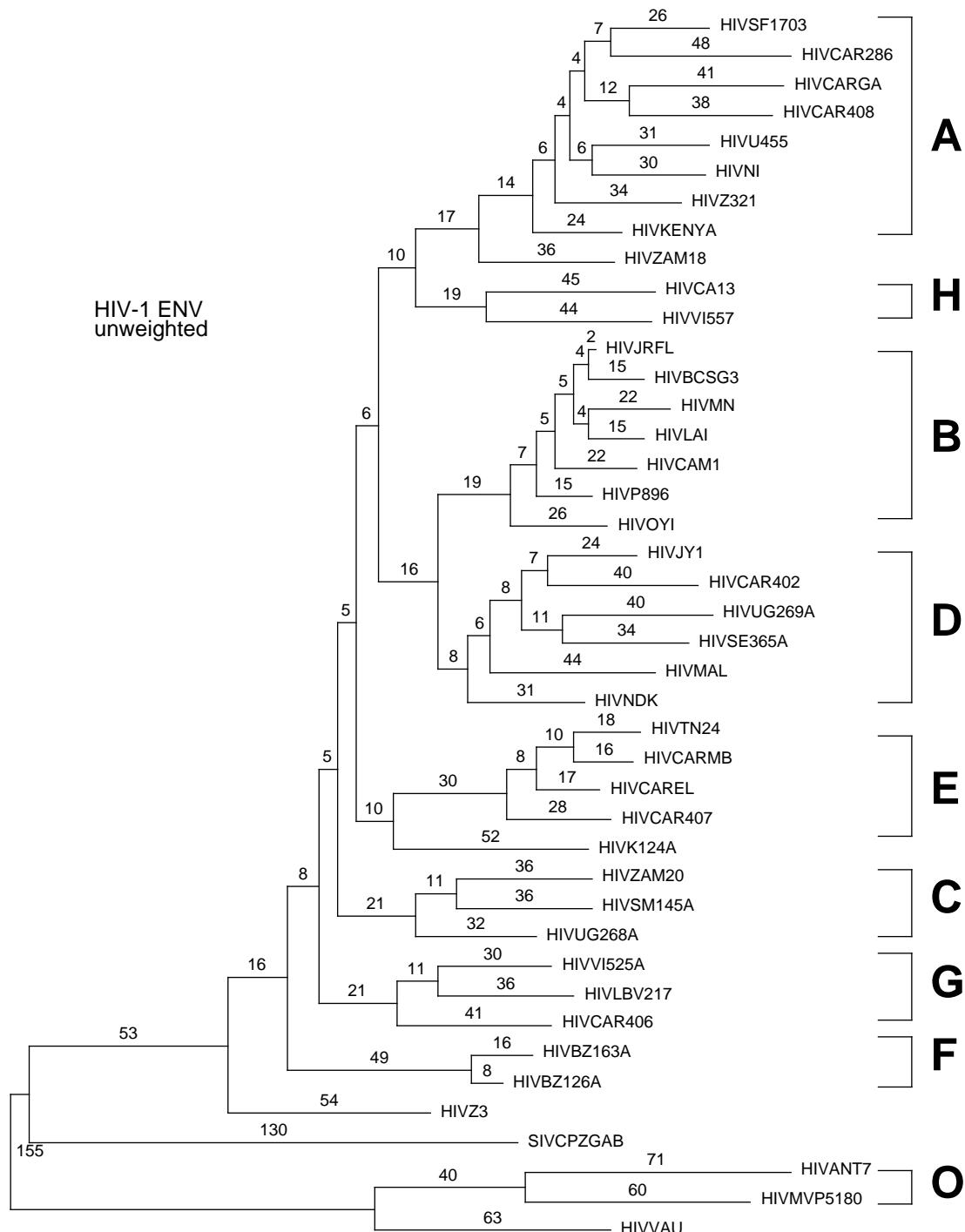
HIV Subtypes



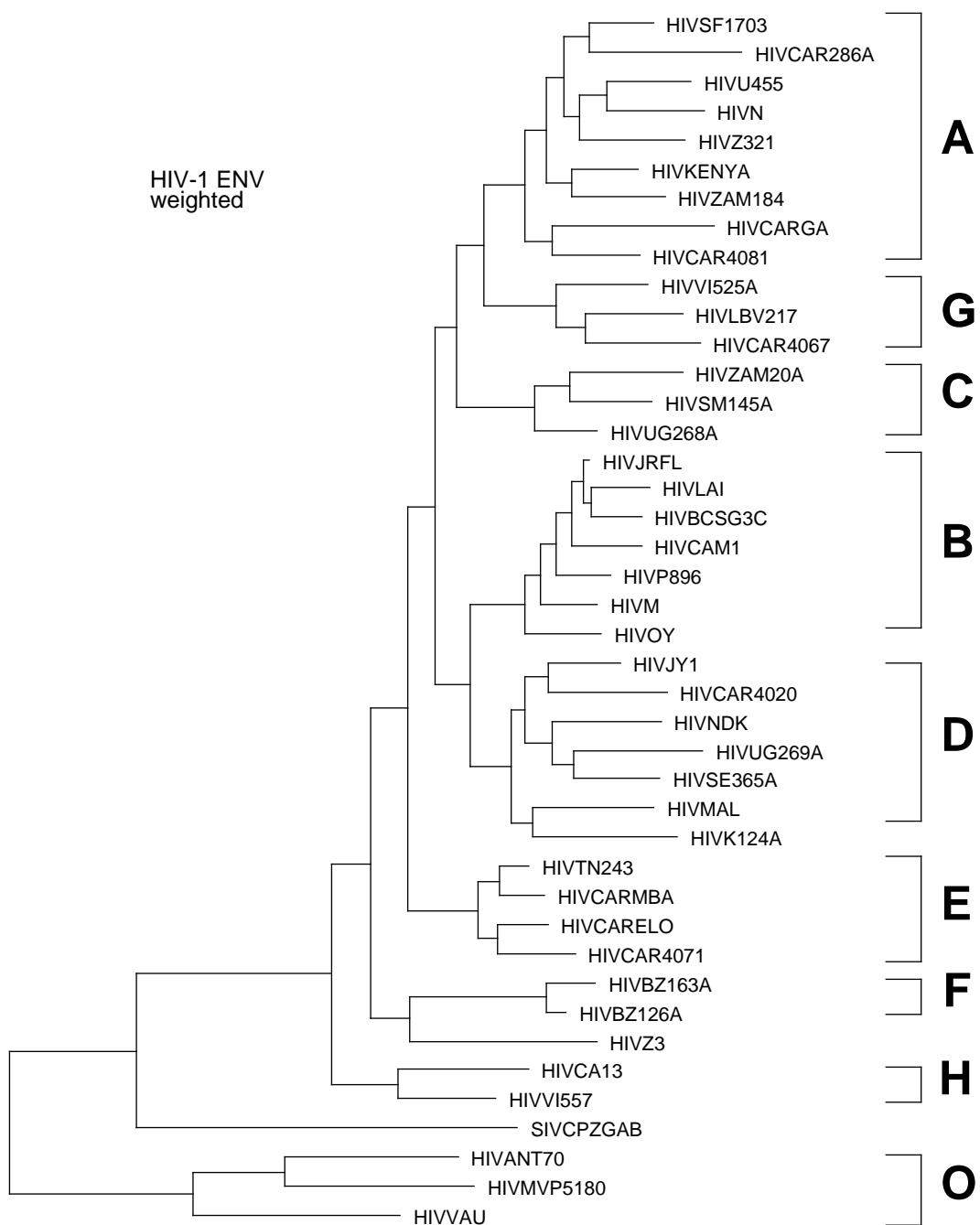
III--24
NOV 95



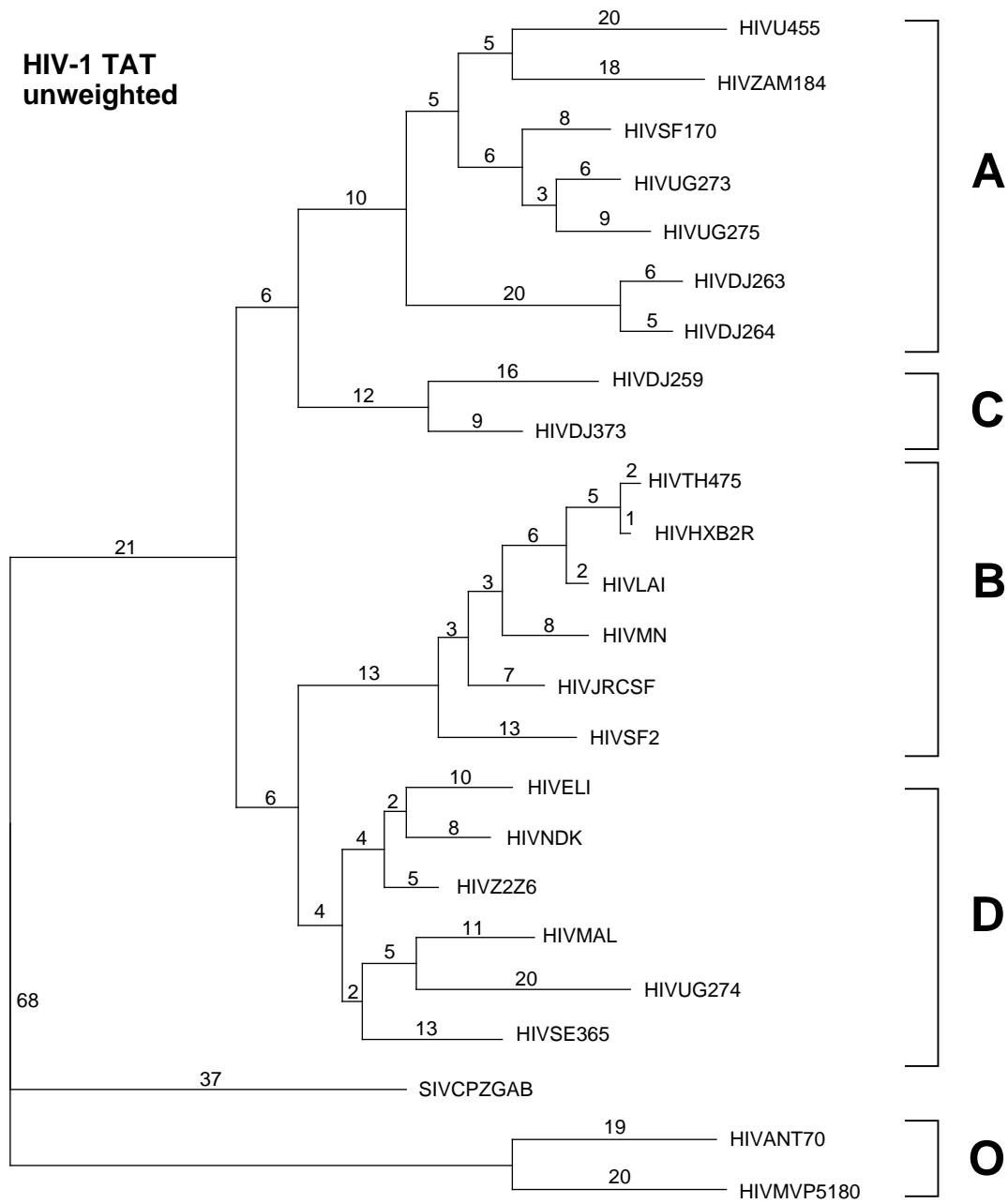
HIV Subtypes



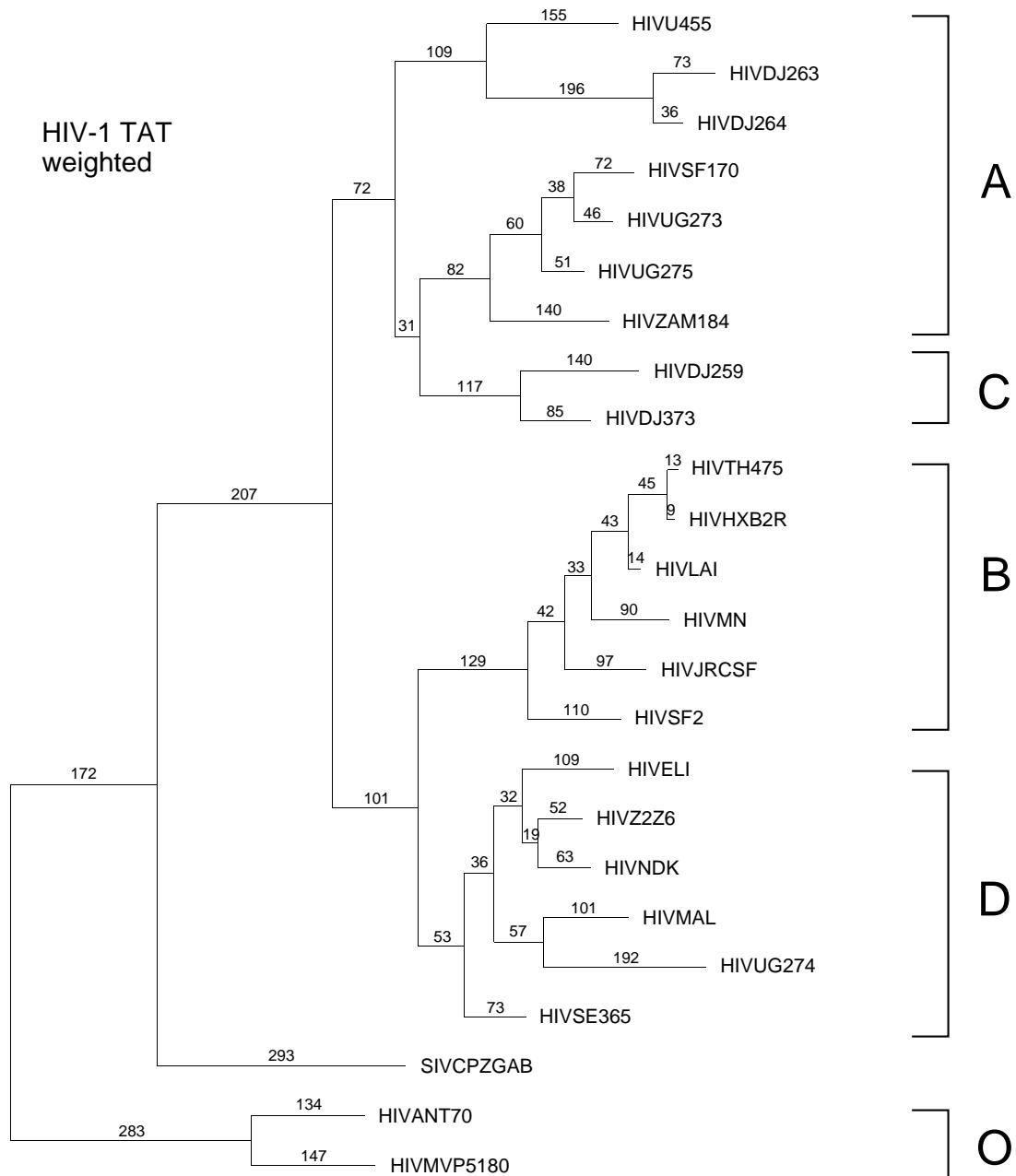
III--26
NOV 95



HIV Subtypes



III--28
NOV 95



This page intentionally left blank.

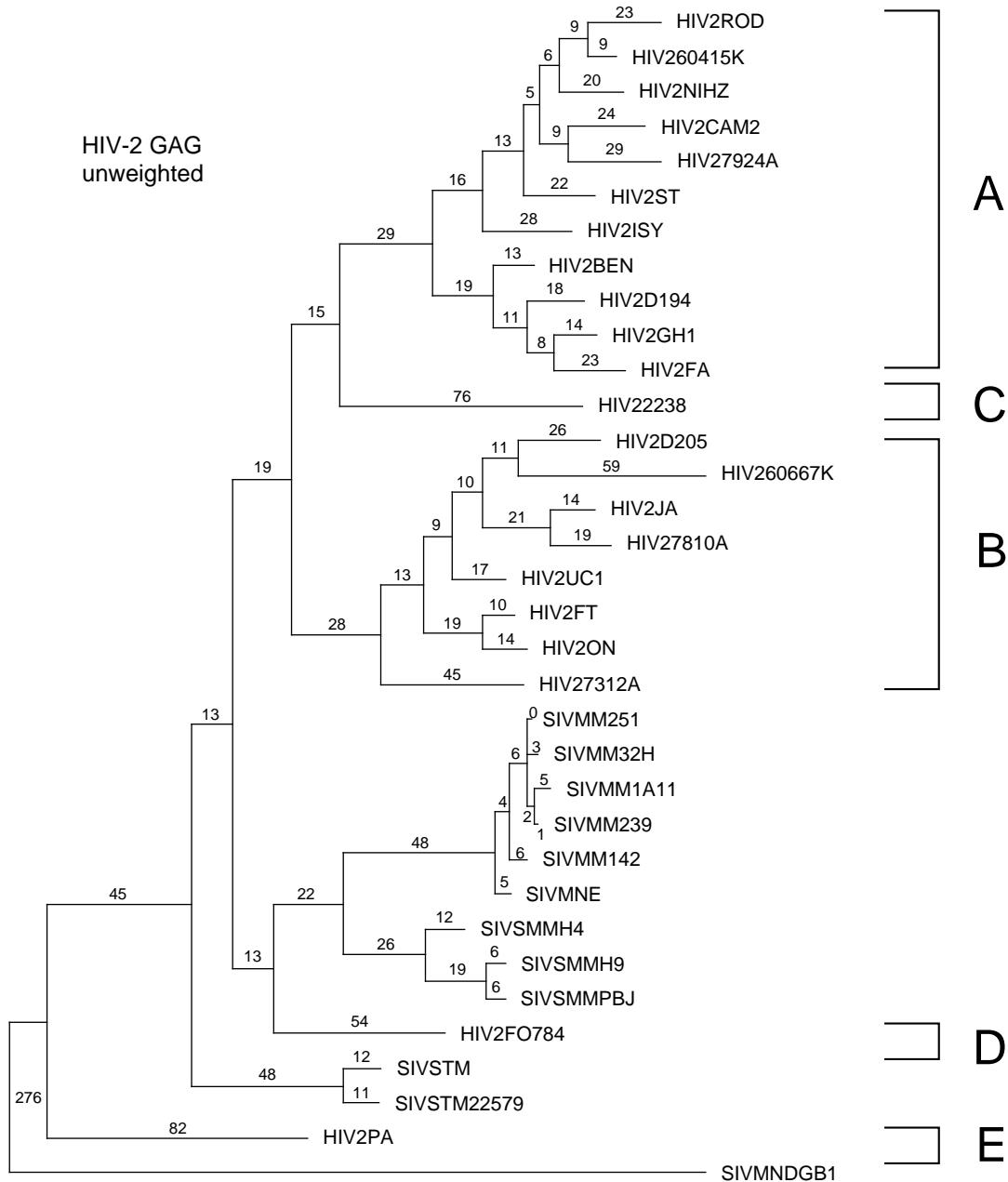
3. HIV-1 Subtype Summary

From various phylogenetic analyses, including those shown on the previous pages of this section, we can update the table of HIV-1 sequence subtype representatives originally presented on page III-2 of APR 94. Not all sequences are listed in the table; this table emphasizes complete *env* and *gag* coding sequences and concordant and discordant *env-gag* pairs. Subtype assignments derived from V3 coding sequences, which in a few instances may not agree with the following assignments, are summarized in III.E. The subtype A sequences are particularly diverse, leading to uncertainties (The designation 'U' indicates uncertainty). Group O sequences have not been subtyped. HIV-2 *gag* subtypes are shown in the following phylogenetic trees, first an unweighted then a weighted tree.

TABLE I. HIV-1 Subtypes

<i>env</i> Subtype	<i>gag</i> Subtype	Sample Locale(s)
A U455, DJ258 DJ264, Z321, SF170, UG031, RW009, RW020, CAR423, CAR4054	A U455, DJ258 K112, K29, K88, VI32, CI20, K7, K98, CI59, LBV2310, CI51, CI44, CI32	Central Africa
A VI191	G VI191	Central Africa
D K124(?), UG266	A K124, UG266	Central Africa
E TN243 (CM243), CAR4017, CAR4071, TH011, TH006	A TN243 (CM243)	Thailand, Central African Republic
B LAI, JRCSF, D31, SF2 TB132, BR014, TH014	B LAI, JRCSF, D31, SF2 TB132, PHI136, PHI153	North America, Europe, South America, Asia
C ZAM18, SM145, UG268 NOF, D747, BR025, DJ373	C ZAM18, SM145, UG268 DJ259, VI131, SE364	Central and Southern Africa, India, Brazil
D ELI, NDK, SE365, UG274 UG038, UG021, CAR4020	D ELI, NDK, SE365, UG274 VI120, G109, K31, G270	Central Africa
F BZ163, BZ126, BR7944	F VI174, BZ162, VI325	South America, Central Africa, Europe
G LBV217 CAR4067, CAR4081	G LBV217	Central Africa, Taiwan Russia
G VI525	H VI525	Central Africa
H VI557 (partial)	H VI557	Central Africa
U ZAM184, Z3	U	Central and Southern Africa
O Group ANT70, MVP5180, VAU	O Group ANT70, MVP5180	West Africa, France

HIV Subtypes



III--32
NOV 95

